IHS Data Warehouse Project Overview

Information Technology Support Center
Albuquerque, NM
November 2003

Project lead:	Stan Griffith
Component leads:	Paul Golis
	Michael Gomez
	Lisa Petrakos

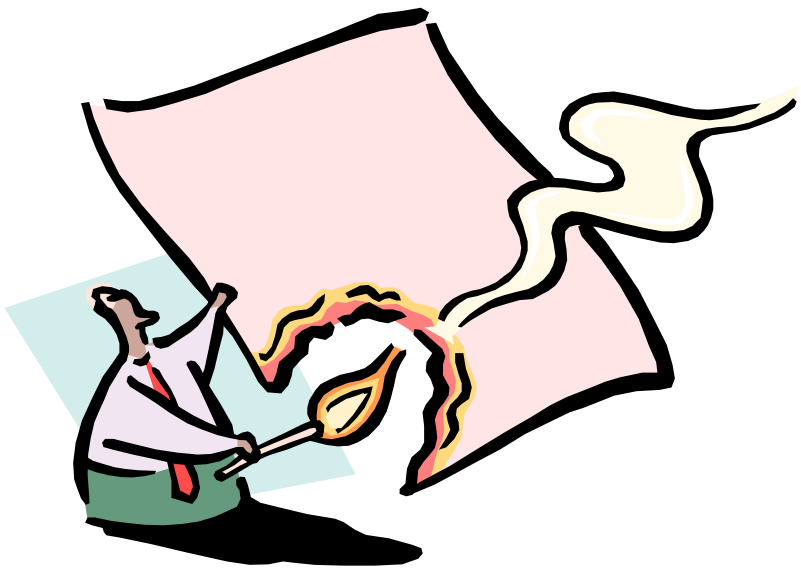
Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

Why Do We Need National Data?

- Requests for information are ever increasing
- These ever increasing requests will likely continue
- But our resources will not increase commensurate to these increased demands

Why Do We Need National Data?



- By deriving information from existing data, data that are already being collected for other purposes,

  we can reduce local data collection burdens and bureaucratic overhead at all levels

Why Do We Need National Data?

Information means money

- GPRA is tied to budget requests
- Funding for special needs
  - Racial disparities
  - Diabetes
- ORYX and JCAHO accreditation

Why Do We Need National Data?

National level data can improve clinical care

- Patient safety

Why Do We Need National Data?

Information provides program direction

- Needs assessment
- Resource allocation
- Outcome performance
- Facility planning

Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

How should we design our national DB?

- We have conflicting needs.

We need more data…

The breadth of information needs is increasing

- Statistical
- ORYX
- Epidemiology
- Diabetes
- Pharmacy
- Cost management
- Bioterrorism

etc.

… in one, all-inclusive repository…

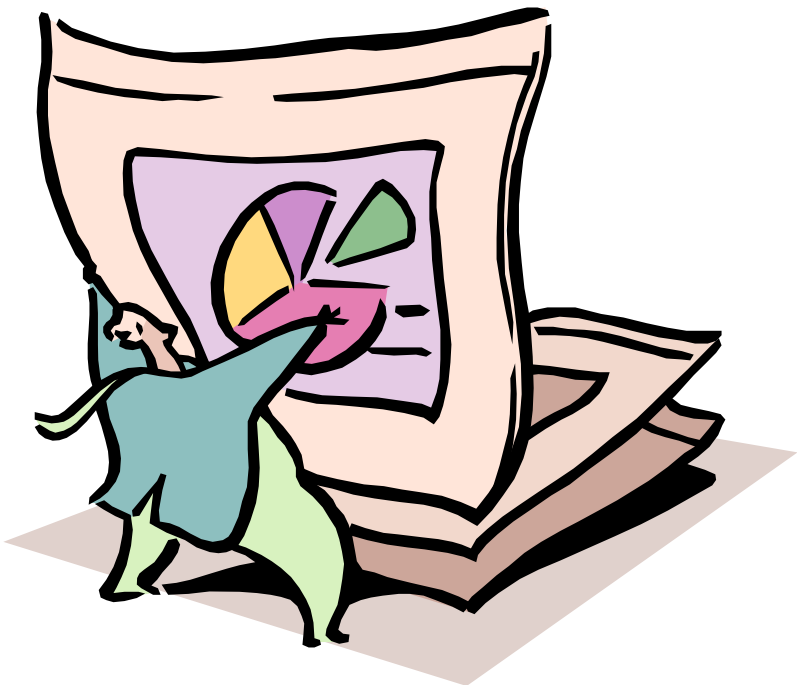
One central collection point for all data

- Verify receipt of data
- Analyze and provide feedback on
  - Timeliness of data
  - Data quality
  - Unexpected deviations from historical norms
- Maintain a "single version of historical truth"
- Maintain all the information content of the data

… but …

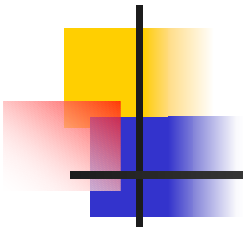
- One large information repository becomes just too unwieldy for reporting

… we need focused, efficient DBs…



We need efficient, user friendly access to data

- Ease of user access
- Search-efficient DB structures
- Subsets of data – just what we need for specific uses
- Transformed data



… still, on the other hand…

But data marts alone just cannot maintain the

- Flexibility
- Granularity
- Scalability
- Power

that will be required to meet all needs now or in the future

… because we have to accommodate the future.

The architecture has to be able to accommodate change, inevitable change.

There will be future needs for information that even the most astute among us cannot now anticipate.

So… we need both!!!

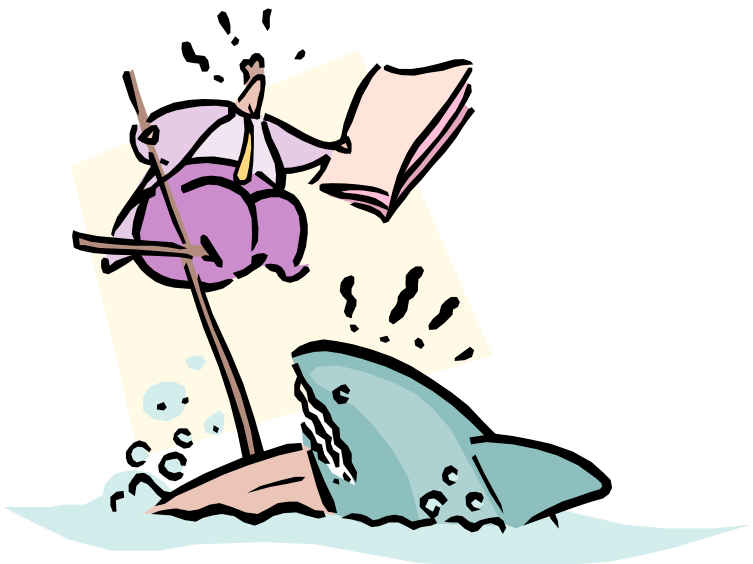
We need one, large, replete, and powerful data warehouse,

that provides information to more focused, user-friendly, efficient data marts.

Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

The Dilemma…

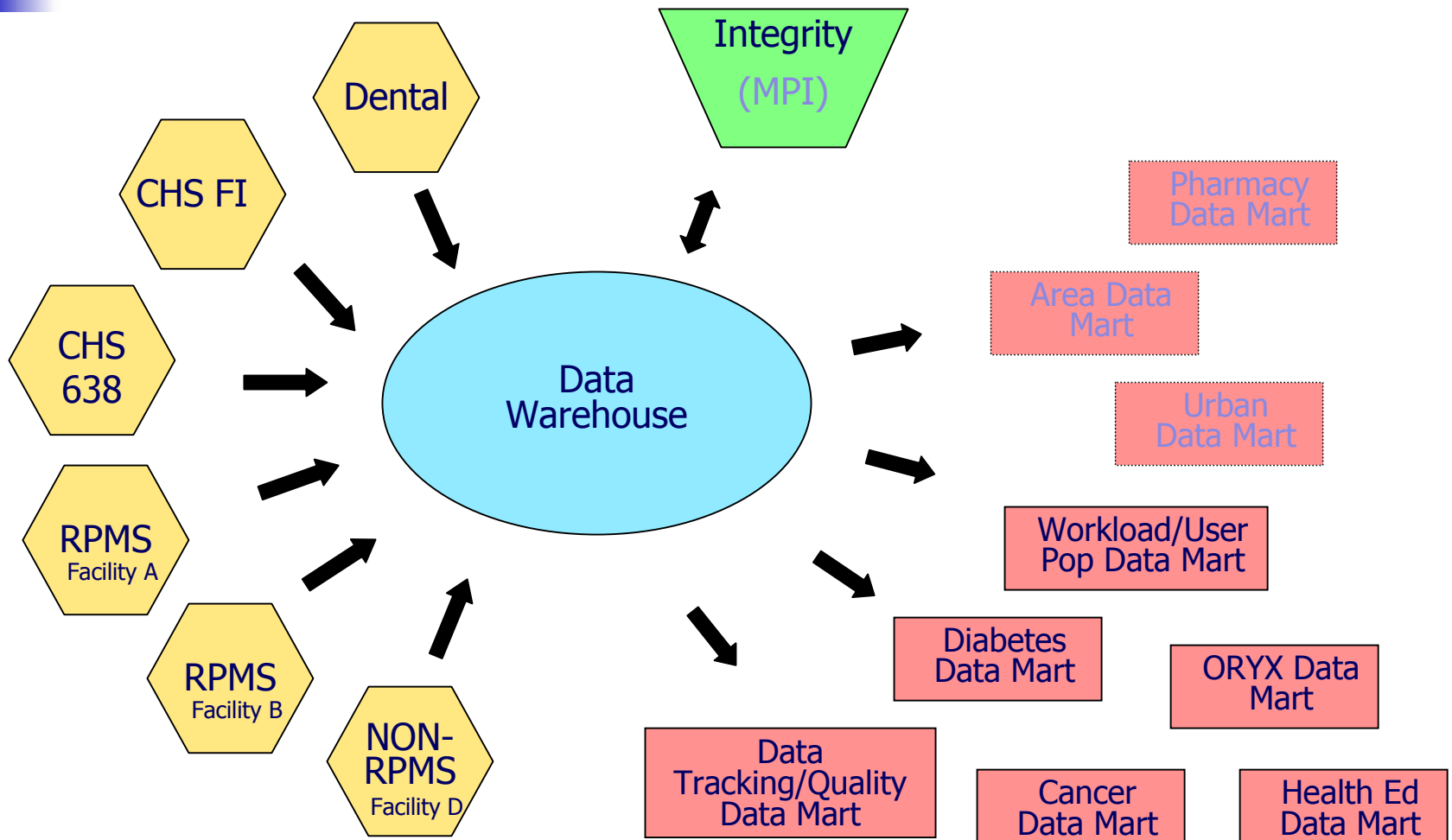


- We need a complete, all-inclusive solution and we needed it yesterday
- But to plan, build, and implement a complete solution would take far, far too long and would not deliver incremental value to our users along the way

The Solution

- Build the data warehouse environment in increments that deliver value to users at each step
- The most critical user needs will be addressed first

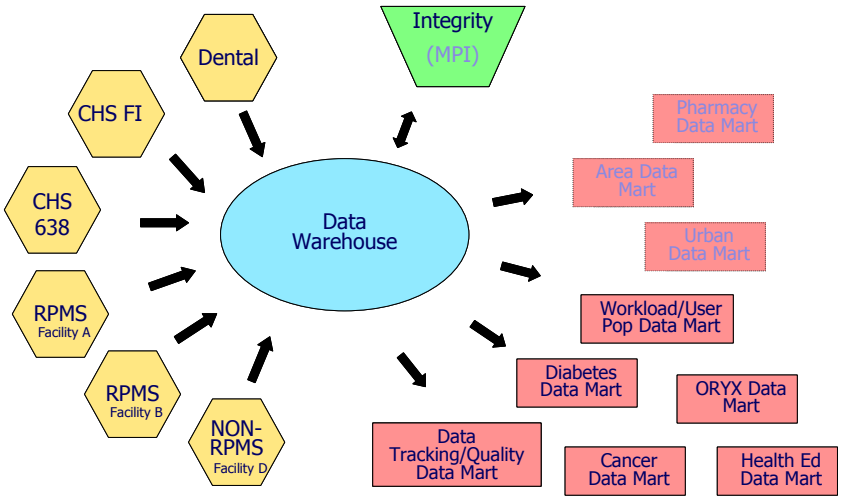
High Level DW1 Architecture



This is a Data Warehouse, not an Operational Data Store

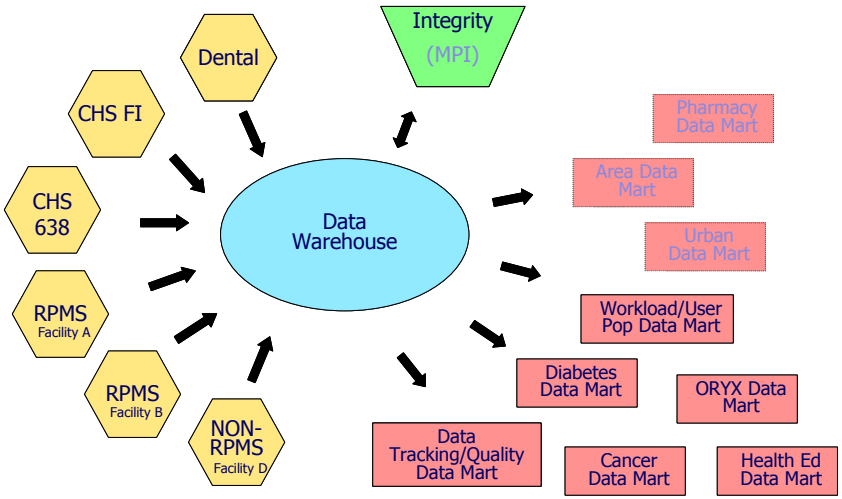
- A data warehouse:
  - Maintains periodic historical snapshots
  - Nonvolatile
  - Time-lagged, periodic exports
  - More focused on retrospective analyses
- An operational data store:
  - Maintains current state
  - Volatile
  - Real time exports
  - More focused on ongoing, up-to-the-minute individual care, billing, etc.

The DW1 Environment Will…



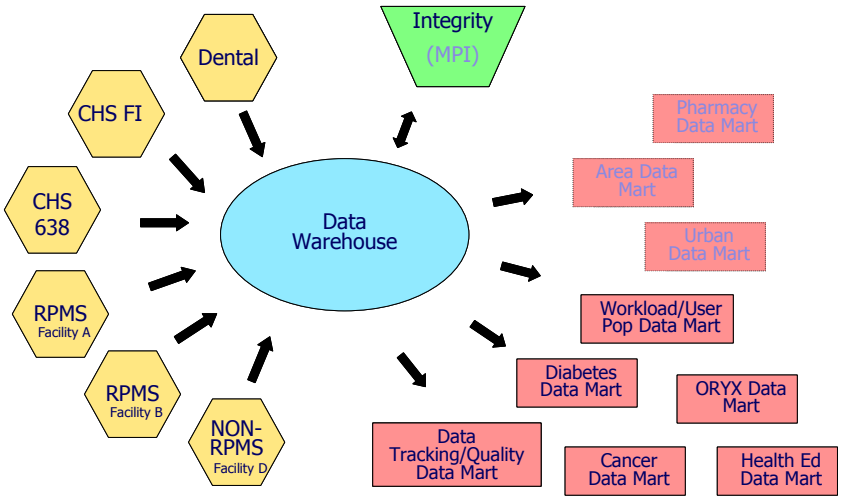
- Accept registration and encounter-based patient data…
- …from sites using RPMS or other IT systems…
- …and CHS data from the Fiscal Intermediary and Area databases.

The DW1 Environment Will…



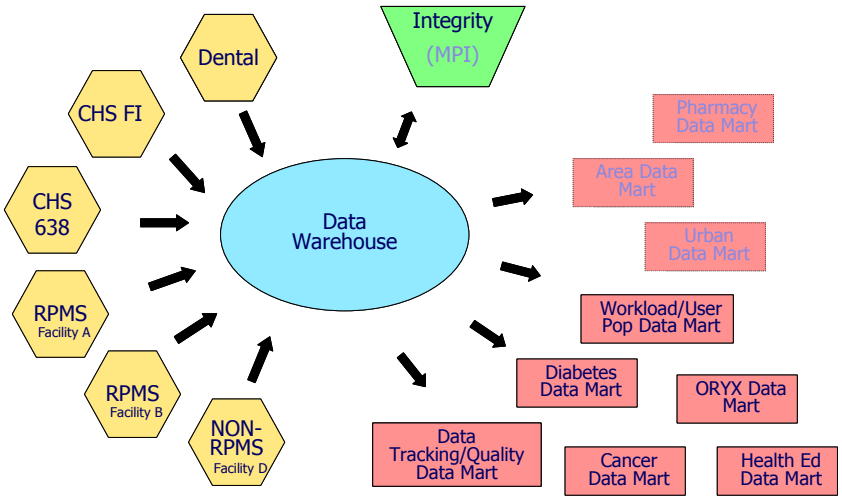
- Collect and store information almost exactly as it is received from the field …
- …maintaining historical snapshots of those data…
- …maintaining as much of its informational content, as possible.

The DW1 Environment Will…



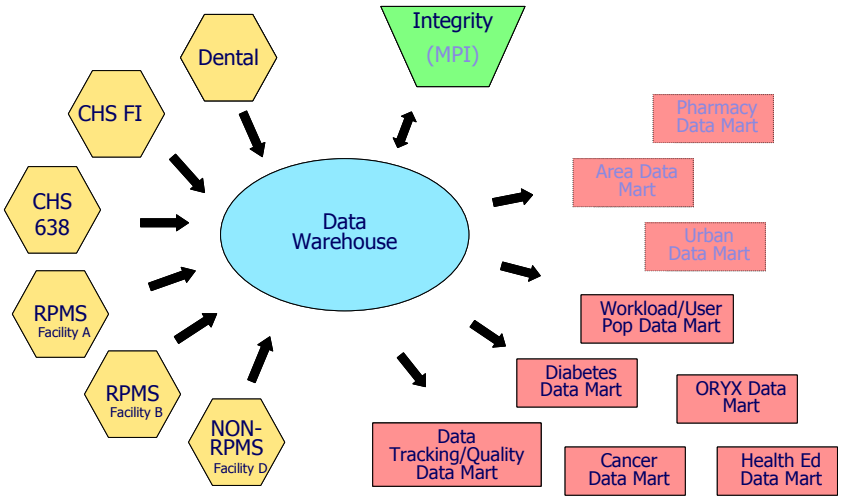
- Provide accurate and helpful information about the data it receives to our field colleagues:
  - Record counts
  - Timeliness of data
  - Less than expected counts based on historical benchmarks
  - Missing data in fields, erroneous codes

The DW1 Environment Will…



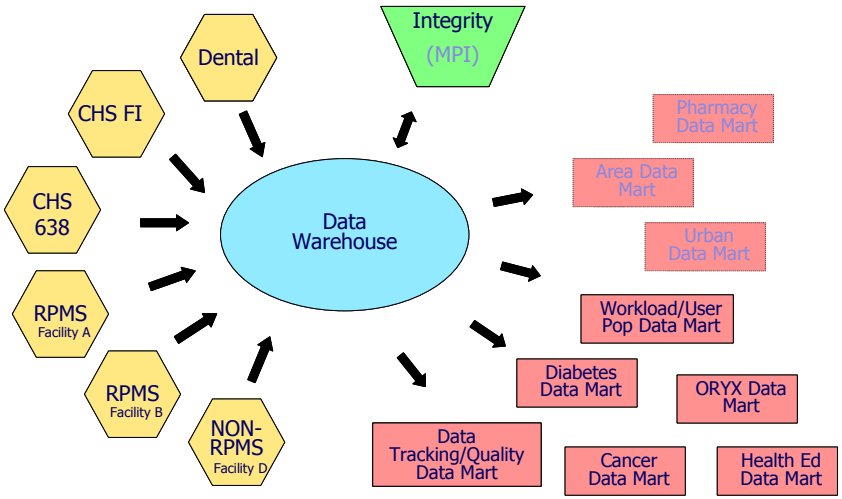
- Supply data to various data marts from which more targeted information can be gathered
  - Workload/User Pop
  - Data Tracking/Quality
  - Outcome measurement (ORYX, GPRA)
  - Organizational units (Area, Urban)
  - Programs (diabetes, cancer, health education, public health nursing, epidemiology, dental, pharmacy)

The DW1 Environment Will…



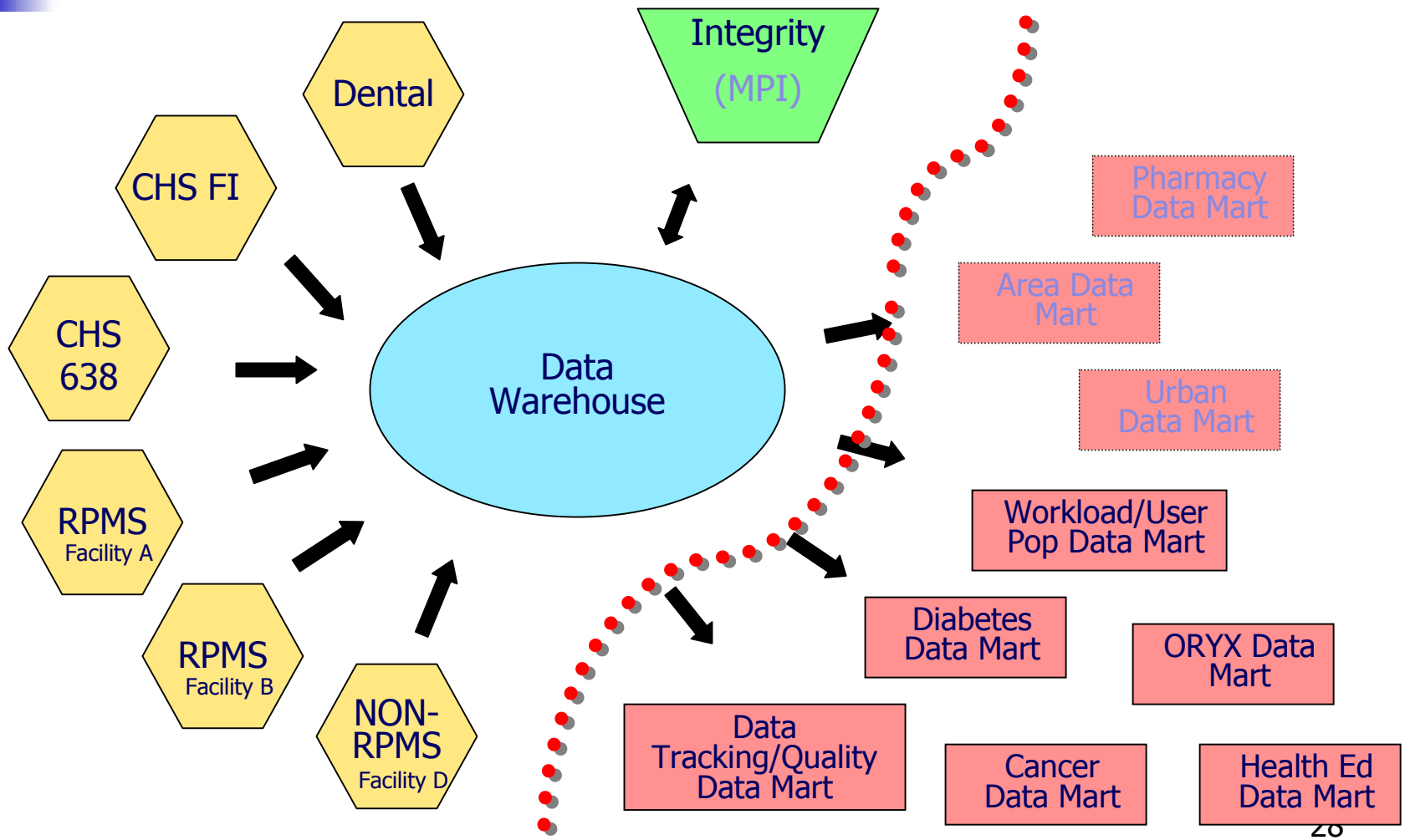
- Use a "probabilistic matching application to resolve by individual across facilities…
- …and use a more complete "Master Person Index" application, when available.

The DW1 Environment Will…



- Employ healthcare industry standards (HL7) for data transport and messaging …
- …employing an Integration engine to route those data from sites to the proper destinations (e.g., the DW, an Area database).

DW Organizational Architecture



DW Organizational Architecture

- Underlying DW database and infrastructure
  - ETL, staging and target tables, administrative and lookup tables
  - Primary responsibility – ITSC
  - Provides data to populate the Data Marts
  - Limited number of "super users" with search access to the DW db
  - Users of this component are the Data Marts

DW Organizational Architecture

- "Component-ized" data marts
  - Data pulls and data mart databases or other structures
  - Primary responsibility – programs, disciplines, organizational units
  - ITSC "hosts" the infrastructure
  - Users of this component are the programs, disciplines, organizational units
  - Provides reports and an ad-hoc search capability to these users

Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

Benefits for our Users

- Improved report accuracy because of improved linking between registration and encounter records
- More confidence in data - better able to verify reports with local data
- Improved data tracking and data quality assessment tools
- Better able to report patient "outcomes"
- Better support for Areas' performance measurements, health status assessments, and surveillance

Benefits for our Users (cont.)

- Better consistency among reports - a rigorously gathered, verified, and maintained "single version of truth"
- Lessened local burdens in producing GPRA measures
- More choices for ORYX users
- More information available for Tribal EPI Centers
- Better information to support Diabetes monies

Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

DW1 Timeline

- Begin data export by December 1, 2003
- Integrated step-wise loading and testing of data over 9-10 months
- Include historical load of data (encounters dated October 1, 2000 to present).
- Run in parallel with current production systems (NPIRS & ORYX) for a year
- Go-no-go decision on new DW mid summer of 2004

Critical Issues for DW Load

- Adequate staffing and hardware for the Interface Engine
- Successful testing of the entire RPMS export process
- Full cooperation and technical support for the load by all Areas and local programs

Why a Comprehensive Re-export?

- Completely redesigned national repository
- Completely redesigned export
- So… data in NPIRS and the DW are not interchangeable. We cannot reuse previously collected data.

What Is So Different?

- Unique registration and encounter record IDs
  - Allow accurate linking of records
  - No consolidation of registration records - full historical record
  - Better able to replicate local level reports
- Change in export allows more complete information
  - All multiple entries such as dental codes, diagnoses, and immunizations
- Ability to store additional clinical fields allowing:
  - More choices for ORYX users
  - Lessened local burdens in producing GPRA measures
  - More information available for Tribal Epi Centers
  - Better information to support Diabetes monies

Why a Comprehensive Re-export?

Above all

All must have confidence that the data is accurate, complete, and verifiable from the beginning

Why a Comprehensive Re-export?

Data quality is paramount.

We cannot afford to just produce a system that duplicates much of what was criticized in the previous system.

Why a Comprehensive Re-export?

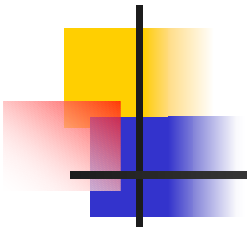
The investment (albeit not insignificant) required to do it right from the start will be much less than what would be required to try and "patch" things later, something that ultimately cannot be successful anyway.

Data Warehouse Project

- Why Do We Need National-level Data?
- Why a DW?
- DW Design
- User Benefits
- DW1 Timeline
- The Future

Future Directions

- Many, many more data marts
- Data exploration and mining warehouses
- Archiving system
- Continuous improvements in the underlying data
- Close monitoring of VHA EHR project
- Increasing incorporation of data standards
- Additional types of data



It is change, continuing change, inevitable change that is the dominant factor in society [health care] today.

No sensible decision can be made any longer without taking into account not only the world as is, but the world as it will be.

Isaac Asimov